

# Behavior of Protein-Inspired Synthetic Random Heteropolymers

Shayna L. Hilburg<sup>†</sup>, Zhiyuan Ruan<sup>‡</sup>, Ting Xu<sup>‡,¶,§,||</sup>, Alfredo Alexander-Katz<sup>\*,†</sup>

<sup>†</sup>Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

<sup>‡</sup>Department of Materials Science & Engineering, <sup>§</sup>Department of Chemistry, and <sup>¶</sup>Tsinghua-Berkeley Shenzhen Institute, University of California, Berkeley, California 94720, USA

<sup>||</sup>Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

---

**ABSTRACT:** Random heteropolymers (RHPs) are an interesting class of materials useful in many theories and applications. While previous studies typically focused on simplified RHP systems, here we explore a more complex scenario inspired by highly heterogeneous molecules like proteins. Our system consists of four monomers mimicking different classes of amino acids. Using Molecular Dynamics simulations and Small-Angle X-Ray Scattering, we explore dynamical and structural features of these RHPs in solution. Our results show the RHPs assemble with heterogeneous interfaces reminiscent of protein surfaces. The polymer backbones appear frozen at room temperature on the nano- to micro-second timescale with molten-globule morphology, albeit their conformational space has multiple metastable conformations for a given sequence, drawing comparison to Intrinsically Disordered Proteins. Local connectivity and chemistry are also shown to have substantial impact on polymer solvation. The work presented here indicates that RHPs share similarities with proteins to be leveraged in bio-mimetic and bio-inspired applications.

---

## INTRODUCTION

Random heteropolymers (RHPs) are polymer chains containing two or more monomers arranged in random sequences.<sup>1</sup> They are easily synthesized and have been used widely in industry for film packaging, lubricants, injection molding, and surface modification.<sup>2–4</sup> More recently, however, it has been shown that RHPs can display some properties thought to only be possible for proteins. For example, using RHPs, researchers have been able to create what are called polymer nanoparticles, which can be thought of as single polymer globules that mimic proteins. These single polymer nanoparticles have been shown to have interesting properties, from drug delivery to catalysis.<sup>5</sup> In fact, in the last few years, a special class of RHPs was shown to perform natural functions such as helping proteins fold properly, stabilizing proteins in non-polar environments, inserting membrane proteins into lipid bilayers, and even forming proton membrane channels,<sup>6–8</sup> which are normally only possible with complex natural proteins. It is important to understand if such RHPs share commonalities with natural proteins. Furthermore, the activity and function of such RHP solutions, which contain a vast amount of the available sequence space in each experiment, have been shown to scale proportionally with concentration, particularly in the low concentration regime.<sup>8</sup> This indicates a general trait of these molecules, rather than a particular sequence being responsible for the observed behavior. Thus, it is of interest to find which features the sequences share. Presumably, there are functions that do

not require specific sequences or folds and that rely mostly on general characteristics such as composition, heterogeneity, and potentially dynamic features. In light of all of these results, there is clearly a need to understand where the observed properties come from, for which it is imperative to obtain a molecular view on the structure and dynamics of such RHPs, as is presented here.

In the past, RHPs were used as model systems to explain some of the properties of proteins, such as the emergence of a single folded state. Theories have shown that protein-like behavior can emerge in RHPs, utilizing simplified models of proteins to show that random copolymers can undergo a transition into a frozen phase from a molten globule phase.<sup>9,10</sup> This transition implied that there are some sequences that have deep energy minima folds, yielding native globular protein-like behavior where a single fold is preferable. Such theories have been extensively studied for copolymers consisting of two monomers (hydrophobic and hydrophilic),<sup>11–14</sup> but have been extended to many monomers types, either in a continuous fashion or in a discrete way such as the work from Pande and co-workers.<sup>9,15–18</sup> Experiments focused on natural amino acid heteropolymers to draw direct corollaries to protein properties were able to confirm many of the theoretical findings.<sup>19–21</sup> Further development of these models, in addition to work that elucidated a funnel-shaped energy landscape, showed that for some sequences, ordered globular protein folding can occur on biologically relevant timescales and lead to a single

stable conformation.<sup>22–24</sup> While these illuminating theories have been very useful in understanding and mimicking native protein folds,<sup>25,26</sup> today we know that specific chemistries might play a particular role and that the strength of the interactions (assumed constant in the models) is modulated by the molecular environment itself.<sup>27–29</sup> Furthermore, in the last fifteen years or so, a large population of natural proteins has been discovered that does not have a precise fold, and therefore, RHPs, and their phases which can present multiple stable states, have regained popularity. In fact, it has been found that approximately half of the eukaryotic proteome contains large regions of disorder with 30–40% being Intrinsically Disordered Proteins (IDPs).<sup>30,31</sup> In such IDPs, the assembled structures are dynamic and variable, making the structure to function relationship far more complex. IDPs typically have lower hydrophobicity and greater net charge than their ordered counterparts.<sup>32</sup> IDP properties influence the dynamics and introduce multiple low energy states, allowing for an increased number of possible interactions and binding mechanisms than proteins with a single more static stable conformation. The wide range of IDP functions in the cell indicates that having a flatter energy landscape with multiple metastable states is beneficial, and at times vital, in enabling biological functions including membraneless organelle formation, some enzymatic reactions, and regulatory processes.<sup>31–36</sup>

On the other hand, while natural RHPs based on amino acids are interesting, there are many other synthetic monomers that can be used, presumably providing functionalities not seen in the natural world. In fact, synthetic chemistry has made significant strides since the original protein folding investigations. The advanced synthetic techniques developed in recent years, such as reversible-deactivation radical polymerization methods, have massively expanded experimental libraries of statistically random heteropolymer systems.<sup>37,38</sup> Polymers which self-assemble into collapsed, soft, nanoscale structures through a variety of transient or covalent crosslinks are actively being studied as analogs to IDPs and referred to as Single-Chain Nanoparticles (SCNPs).<sup>39,40</sup> A wide variety of chemistries are employed for these studies and, while some research is being conducted with native amino acids, the SCNP field now focuses on more easily synthesized building blocks that contain hydrophobic and hydrophilic components. SCNPs are of paramount interest for their potential applications leveraging antibacterial properties, performing chemical encapsulation, as drug delivery vehicles, and as enzyme mimics incorporating catalytic centers.<sup>5,41–47</sup>

Nevertheless, SCNP research, like many of the theoretical studies preceding it, has delved mostly into two-component RHP systems. While theory work often studied systems of hydrophobic and hydrophilic units of uniform size, SCNP experimental work naturally contains a

greater variety of chemical and steric heterogeneity, even when limited to only two monomer species. Several popular methods in SCNP literature rely on methacrylate-based monomer building blocks due to the variety of available synthesis and functionalization methods, as well as inherent amphiphilicity of these constituents.<sup>5,40,42,48–52</sup> These methacrylate-based heteropolymers have been used to demonstrate the dramatic effect side chain chemistry can have on polymer collapse and resulting nanoparticle behavior, as revealed through common experimental SCNP analysis techniques.<sup>48,53</sup>

Inspired by the recent functionalities observed in RHPs,<sup>6–8</sup> here we study the molecular structure and dynamics of complex RHPs containing up to four different monomers using molecular dynamics as well as X-ray scattering. In general, our results show that such RHPs fold into globular states with frozen backbones, in agreement with the theories of RHPs, yet our studies provide molecular level details of chain organization, packing, and solvation which are important for understanding the observed functionalities of these copolymers. Furthermore, we find some interesting effects such as the fact that the backbone mobility decreases as one increases from two components to four components, even if the added components are charged, and by themselves swell the globule. In fact, we find that there is no obvious correlation between density and backbone mobility in these systems. Such results are beyond what is attainable with previous theories due to the intrinsic assumptions they rely on. Overall, our work provides the first full molecular characterization of these synthetic macromolecules and supports previous studies relating RHPs to proteins, since RHPs assemble into many different metastable states as IDPs do and their structural characteristics resemble those of proteins. Furthermore, the assembled states are stable over time scales relevant for many functions such as binding. Our work, thus, reveals how RHPs create bio-inspired globules in water and has important ramifications for a wide range of systems including native sequence defined biomolecules and synthetic biomimics.

## METHODS

**Computational Methods.** Random heteropolymer sequences with a degree of polymerization of 100 were simulated using the Compositional Drift program targeting an average of 50:25:20:5 ratios of MMA:OEGMA:EHMA:SPMA, resulting in 9216 sequences.<sup>54</sup> The OEGMA used in simulation contains 9 poly(ethylene glycol) (PEG) repeat units terminated with a methyl group corresponding to the experimental OEGMA repeat units with initial monomer molecular weights of 500 g/mol. Ten sequences were then randomly selected from the output and randomly assigned chirality for a racemic mixture. Each sequence was then simulated according to the following protocol.

Monomers were assigned partial charges from a quantum mechanics calculation performed using Gaussian 03 Revision D.01.<sup>55</sup> AmberTools17 was then employed to parameterize the polymers and assemble sequences into extended chains to be simulated using Amber16.<sup>56</sup> The General Amber Force Field was used with 2 fs time-step and Langevin thermostat for all simulations. When pressure is specified, it was

maintained using a Berendsen barostat. After a short minimization, an initial equilibration at 300 K was performed for 20 ns and then an anneal cycle was completed using the Generalized Born/Surface Area (GBSA) implicit solvent model. The annealing protocol in implicit solvent heated the system to 500 K in 10 ns and then brought it back to 300 K through five smaller heat and cool cycles over the course of 90 ns, and finally held for 40 ns at 300 K. This anneal was successively repeated for a total of ten annealing cycles. The resulting conformations from each of the ten anneals were then extracted and solvated explicitly using an SPC/E water model with potassium counterions to neutralize the system (in equal proportion to the anionic SPMA monomer) using monovalent ion parameters from Joung and Cheatham into a truncated octahedral geometry box with periodic boundary conditions and 16 Å minimum molecule-to-box distances.<sup>57</sup> The ten conformations for each sequence were then annealed in the explicit water conformations. This protocol included a short minimization, a 2 ns simulation at 1.0 bar, and then a constant volume anneal cycle. This cycle increased temperature to 650 K in 5 ns, held for 25 ns, and then decreased back to 300 K over 40 ns. The system was then held at 300 K for 20 ns at a pressure of 1.0 bar. Each system was then run for a final 40 ns at 300 K and 1.0 bar, which is the portion of the simulation used for equilibrium analysis. Two-component control sequences followed the same protocol, except contain ten runs of one randomly selected sequence of the desired composition rather than ten sequences selected from the Compositional Drift program output, as was the case for the four-component RHP.

Analysis was performed using a combination of AmberTools17 and Python tools.<sup>56,58,59</sup> Radius of gyration is calculated for each conformation as an average over 40 ns including all polymer atoms. Density is calculated using the mass of the sequence and an estimated volume based on convex hull analysis on the mean structure including all polymer atoms for each conformation. This is likely a low estimate of density due to the use of convex hull volume which could incorporate voids near extended chains at the surface of the structure. The absolute value of dihedral angle means and standard deviations are based on the distribution of each dihedral angle in the polymer backbone over 40 ns for each conformation. Alpha shape surface analysis uses Delaunay triangulation of the mean structure using residue positions as points, similar to methods for protein analysis but without incorporating the van der Waals radii.<sup>60,61</sup> To assess the surface by monomer, an alpha value of 0.075 is used for analysis of the C<sub>m</sub> atoms (the first side-chain carbon attached to the carboxyl oxygen as identified in Supplementary Figure 3), corresponding to a maximum circumradius of 13.3 Å. Alpha of 0.3 corresponding to maximum circumradius of 3.3 Å is used in the same analysis including all heavy (non-hydrogen) polymer atoms. Values reported for the aforementioned properties are the means and standard deviations over the one-hundred conformations resulting from running ten conformations for each of the ten sequences.

Root-mean-square fluctuations (RMSF) were calculated for each monomer in the one-hundred conformations. Translational and rotational motion were first removed using root-mean-square deviation (RMSD) of the backbone with the mean structure as reference. Backbone RMSF was then calculated from the  $\alpha$ -carbon atom in each repeat unit while side-chain fluctuations were characterized from tail atoms, using the terminal carbon atom in MMA, OEGMA, and EHMA side-chains and from the sulfur atom in SPMA. These atoms identifications are depicted in Supplementary Figure 3. Reported RMSF values include the mean and standard deviation between the ten-thousand values found from analyzing each residue in each conformation over 40 ns for the equilibrium values and in 5 ns intervals for RMSF as a function of temperature during the anneal.

The population of water molecules in the first solvation layer was calculated separately for each monomer and averaged over the 40 ns. The values were then normalized by values obtained through reference trimer simulations of each monomer type using MMA-X-MMA where X is replaced by the monomer of interest (MMA, OEGMA, EHMA, or SPMA). Trimer reference simulations underwent a short minimization and then equilibrated for 5 ns and followed by a 100 ns simulation at 300 K and 1.0 bar, from which the maximum number of water molecules in the first solvation layer was selected. Once normalized to become the Normalized First Solvation Shell (NFSS), the middle 90

monomers for each of the one-hundred configurations were selected and used in further analysis. Normalization factors for Solvent Accessible Surface Area (SASA) were also calculated for each monomer using the same reference simulations, and used to calculate Relative Solvent Accessibilities (RSA). SASA was found for each monomer, including all atoms, using the LCPO method implemented in AmberTools, and averaged over the 40 ns. Similarly to NFSS, the middle 90 monomers for each of the one-hundred configurations were selected and used for sequence RSA analysis. All molecular renderings were created using VMD.<sup>62</sup>

**Materials.** Chemicals were purchased from Sigma Aldrich Chemical Co. or Fisher Scientific International, Inc. unless otherwise noted. Water was purified by a Milli-Q water filtration station (18.2  $\Omega$ m cm) before use. Azobisisobutyronitrile (AIBN) was recrystallized from ethanol before use. Inhibitors were removed using cryodistillation (methyl methacrylate, 2-ethylhexyl methacrylate) or by passing through a short column of neutral alumina (ethylene glycol methyl ether methacrylate (Mn ~500 g/mol). 3-sulfopropyl methacrylate potassium salt (98%) prior to polymerization. Ethyl-2(phenylcarbanothioylthio)-2-phenylacetate, Trioxane (TCI, internal standard for <sup>1</sup>H NMR analysis), and Dimethylformamide (DMF, solvent) were used without further purification.

**Polymer Synthesis and Characterization.** The synthesis of RHPs was carried out using reversible addition-fragmentation chain-transfer polymerization (RAFT) as our previous work.<sup>6</sup> Polymerization solutions were prepared by mixing the requisite amounts of reagents and added into 25 mL glass Schlenk. The solutions were subject to 3 freeze-pump-thaw cycles before the Schlenk was placed into the oven. The polymerizations were held at 70 °C for typically 4-8 hours. Global monomer conversion was determined by <sup>1</sup>H-NMR on crude reaction mixtures in DMSO-d<sub>6</sub> with trioxane as an internal standard. The polymerization solution was then precipitated by dropwise addition to rapidly stirring pentane. The resultant polymer was then dissolved in water and transferred to a 2,000 MWCO dialysis tubing, and dialyzed against distilled water for three days. The purified polymer was then lyophilized to remove solvents.

Molecular weight distribution curves, number-average (Mn) and weight-average (Mw) molar mass and dispersity ( $\bar{D} = Mw/Mn$ ) of the polymers were measured by gel permeation chromatography (GPC) using an Agilent 1260 Infinity series instrument outfitted with 2 Agilent PolyPore columns (300  $\times$  7.5 mm). DMF with 0.05 M LiBr was used as eluent at 0.7 mL/min at 50 °C. The columns were calibrated against Poly(ethylene glycol) standards. Analyte samples at 2 mg/mL were filtered through 0.2  $\mu$ m polytetrafluoroethylene (PTFE) membranes (VWR) before injection (20  $\mu$ L). <sup>1</sup>H NMR spectra were recorded in DMSO-d<sub>6</sub> on a Bruker Avance 400 spectrometer (400 MHz) using a 5 mm Z-gradient BBO probe or a Bruker Avance AV 600 spectrometer (600 MHz) using a Z-gradient Triple Broad Band Inverse detection probe.

Small-Angle X-Ray Scattering (SAXS). SAXS was carried out at beamline 8-ID-E at the Advanced Photon Source, Argonne National Laboratory. Samples were dissolved in water at a range of concentrations, from 0.2 to 2 wt%. Samples were measured in 2mm boron-rich thin-walled capillary tubes and subject to multiple short exposures (5 s for each time). 2D scattering results were azimuthally averaged to produce 1D SAXS profiles. Superimposable profiles were averaged and then subtracted from the background. The radius of gyration was obtained from a Guinier analysis shown in Supplementary Figure 4.

## RESULTS & DISCUSSION

Our system examines RHPs containing methyl methacrylate (MMA), oligo(ethylene glycol) methacrylate (OEGMA), 2-ethylhexyl methacrylate (EHMA), and 3-sulfopropyl methacrylate (SPMA) monomers, for which the chemistry is depicted in Figure 1A. Each of these monomers is selected to mimic charged (SPMA), polar (OEGMA), and hydrophobic (MMA and EHMA) properties of amino acids. Ten sequences were selected to

represent a wide range of the statistically distributed compositions from the pool of nearly ten-thousand sequences created through the Compositional Drift program.<sup>54</sup> Each chain is one-hundred monomers in length and is schematically presented in Figure 1C. The molecular weight of each of the selected chains are 22.6 kDa on average, ranging from 17.8 to 24.4 kDa, with mean monomer ratios presented in Figure 1B and compared to the overall program output population in Supplementary Figure 1. By number of each monomer type, the MMA contributes most significantly to the polymer composition; while by mass, or total molecular weight of each monomer type, the OEGMA is the predominant component due to the size discrepancies of the building blocks. Experimentally, GPC of the RHPs used in SAXS analysis indicates a number average molecular weight of 19 kDa with dispersity of 1.4.

While some computational SCNP research has been performed using implicit solvent,<sup>63</sup> we have opted for a fully atomistic representation of our RHPs and the solvent. Through these atomistic simulations with explicit solvent, we are able to sample not only the two-body interactions, but also the multi-body effects which may play important roles in the hydrophobic effect, and solvent effects which cannot be replaced fully by effective interaction energies.<sup>27,64,65</sup> Additionally, atomistic detail for the RHP means that intra-monomeric degrees of freedom are incorporated, therefore accounting for configurational entropy which may be significant in components such as the OEGMA, a far larger monomer than native amino acids. The results of these atomistic MD

simulations showed that all sequences collapse into globules in water through annealing over time, akin to the representative snapshots in Figure 2 (for a full overview of resulting globular structures, see also Supplementary Figure 2).

The average radius of gyration for the resulting heteropolymer globules in simulation is  $1.74 \pm 0.06$  nm, indicating a significant compaction from the initial extended conformations given the heteropolymer contour length of approximately 30 nm. For a two-component system of poly(MMA-r-OEGMA) in similar proportions and degree of polymerization to our RHPs, Imai et al. used SEC-MALLS to identify hydrodynamic radii of 3.7 nm,<sup>51</sup> which, given the exclusion of a strongly hydrophobic monomer like EHMA, corresponds favorably with our simulation results. Studies on related two-component polymers by Guazzelli et al. also showed compaction to a hydrodynamic radius and radius of gyration of approximately 2 nm found through both DLS and MD simulation, though the chemical makeup of their structure was more distinct from ours.<sup>49</sup> Our own SAXS studies, using the four-component RHP, yield a radius of gyration of  $3.44 \pm 0.13$  nm as detailed in Supplemental Figure 4. While this experimental radius of gyration value is significantly larger than the result from simulation, preliminary experimental studies suggest oligomerization of our RHP at this concentration range, which is consistent with prior findings indicating the tendency of RHPs to oligomerize.<sup>19</sup> The Kratky plot in Figure 3, also obtained from SAXS, is used to qualitatively assess

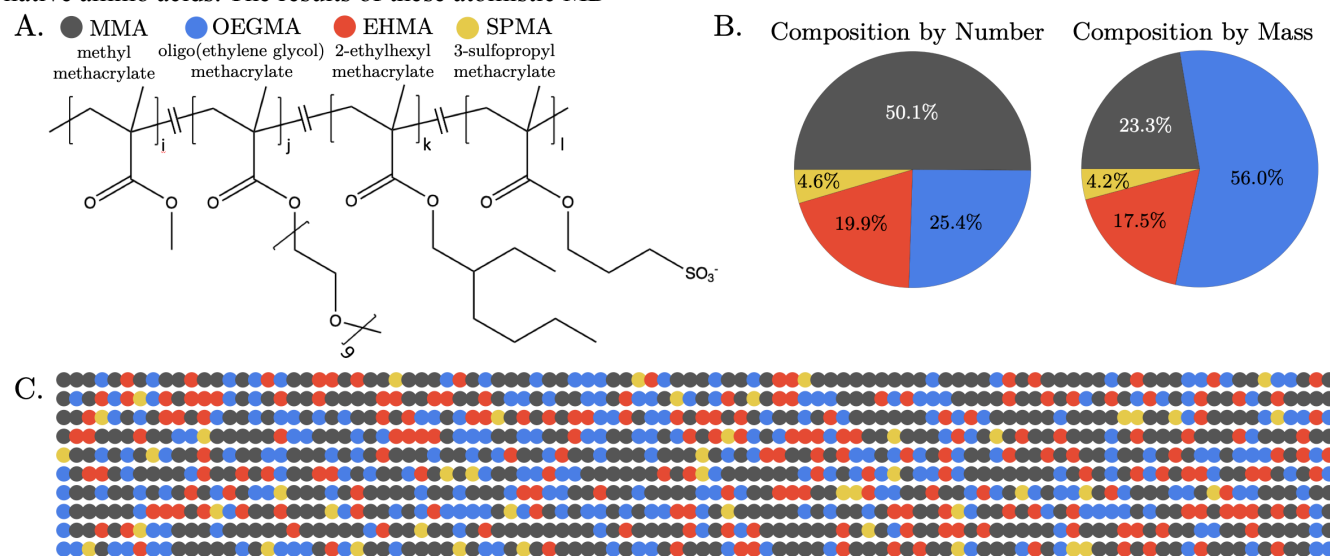


Figure 1. Design of the random heteropolymer system. (A) Polymer chemical structure and color coding for MMA, OEGMA, EHMA, and SPMA. (B) Composition of the heterogeneous polymers studied here and based on previously reported chemistry.<sup>6</sup> (C) Schematic of the ten different sequences selected for simulation.

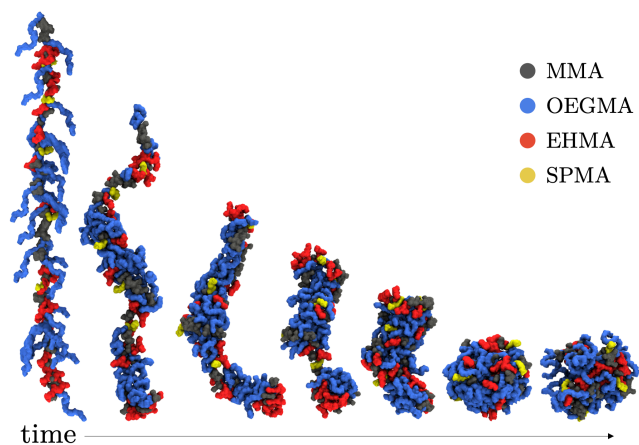


Figure 2. RHP Collapse. Representative collapse of a random heteropolymer in water with residues identified by color with MMA in gray, OEGMA in blue, EHMA in red, and SPMA in yellow. This particular sequence corresponds to sequence 8. The left six conformations are from the first implicitly solvated equilibration while the rightmost conformation is after the explicitly solvated conformation has been equilibrated. Water molecules are omitted for clarity.

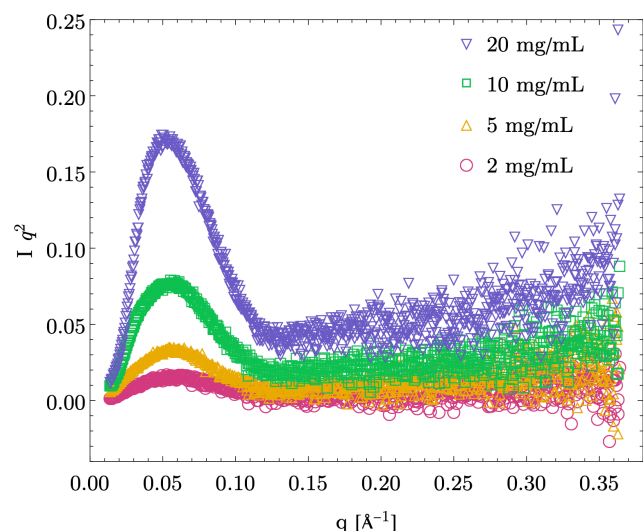


Figure 3. Kratky Plot. Calculated from SAXS data of 19 kDa random heteropolymer at 20, 10, 5, and 2 mg/mL concentrations. The characteristic bell shape of the curves indicates compact structures.

compactness. Our results show a distinct bell shape indicative of a compact or folded structure, consistent with simulation results, since the signature of decreasing  $Iq^2$  at intermediate  $q$  values would not be seen for a random coil.<sup>66,67</sup> These results show that earlier studies experimentally confirming the molecular compactness seen in amino acid based RHP assemblies apply to our synthetic RHP molecules as well.<sup>68</sup>

The mean density approximated from simulation is  $1.10 \pm 0.08 \text{ g/cm}^3$ . This is comparable to the density of bulk poly(methyl methacrylate) reported to be  $1.18\text{--}1.19 \text{ g/cm}^3$ .<sup>69,70</sup> Given the solubility limit of  $0.6\text{--}0.7 \text{ g/mL}$  for poly(ethylene glycol) in water, our RHPs are at nearly double this density and therefore even the hydrophilic

OEGMA side-chains contribute to the globular assembly of the molecules due to the density enforced by covalent connection to the backbone.<sup>71</sup> Visual inspection of each conformation in simulation shows a globule, albeit often aspherical, with heterogeneous surfaces decorated by a significant proportion of OEGMA side chains.

While dense globule formation is an important property of these heteropolymers for future utility in nanoscale delivery or as protein mimics, perhaps the most intriguing aspects of this polymer are the observed dynamics within that globule. Pioneering work by Shakhnovich et al. predicted molten globule formation for RHPs.<sup>72</sup> In this morphology, the backbone is frozen with some side chain motion allowed.<sup>73</sup> Their theories and some experimental work also indicated that side chain motion can vary depending on its degrees of freedom.<sup>74,75</sup> We observe this frozen backbone for our RHPs as predicted by theory, and leverage MD data to examine the heterogeneous dynamics at play with an unprecedented level of detail from the simulations. To access the changes in backbone and side chain mobility, we probed the root-mean-square fluctuations (RMSF) through the annealing cycle during cooling of the annealed RHP back to 300K. We find that the backbone atoms within the molecules respond nearly instantaneously to changes in temperature. The heteropolymer in water is an extremely overdamped system, with its high viscosity likely coming from the small size of the area of interest and the

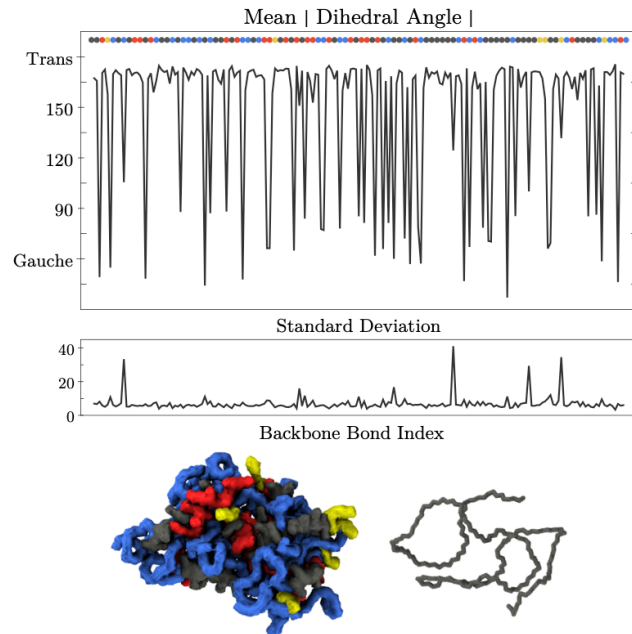


Figure 4. Dihedral Angles. Representative data for the absolute value of the RHP dihedral angle. Mean and standard deviation of the absolute value are shown for each dihedral angle in the backbone over 40 ns with a snapshot of the (Left) equilibrated structure and its (Right) backbone during the simulation with water molecules omitted for clarity. Data shown is for sequence 3 conformation 1, and data for all sequence 3 conformations can be seen in Supplementary Figure 5.



strong intramolecular forces between molecular constituents. Upon equilibration, backbone fluctuations are all quite low, with an overall mean value less than 1 Å, indicating minimal mobility. Assessment of the dihedral angles upon equilibration further support the notion of a frozen backbone. When considering data over the final 40 ns, only  $1.5 \pm 1.0$  % of the absolute value of dihedral angles have a standard deviation greater than 20 Degrees, indicating that the vast majority of trans and gauche conformations remain stable throughout. Representative data of the absolute value of the dihedral angles is shown for a single conformation of sequence 3 in Figure 4, where 4 of the 197 calculated angles show standard deviations above 20 Degrees. In comparison to the  $1.5 \pm 1.0$  % over 40 ns at 300 K, at 650 K,  $86.3 \pm 3.8$  % of dihedral angles have a standard deviation greater than 20 Degrees in just 25 ns. This indicates that the annealing process effectively remodeled the globule and that the equilibrium structures at 300 K are stable. For proteins, the reports of IDP relaxation timescales range from under a nanosecond to several microseconds for various molecular components, and full conformational interconversion likely takes tens of nanoseconds to microseconds.<sup>35,76,77</sup> Globular ordered proteins exhibit millisecond to minute folding timescales. Structurally, our RHPs also have surface to volume ratios that fall in-between globular ordered proteins and more extended IDPs. At a mean of 22.6 kDa, our heteropolymers have total solvent accessible surface area (SASA) values of  $135 \pm 11$  nm<sup>2</sup>, while Gallat et al. found the 43 kDa folded maltose-binding protein to have a SASA of 150 nm<sup>2</sup> and the 45 kDa disordered Isoform Tau-F to have SASA of 470 nm<sup>2</sup>.<sup>78</sup> With this information, we can see that our RHP is more dense and has slower remodeling than many IDPs, although is likely still more dynamic than most globular ordered proteins which configure slowly and exhibit greater compaction. In the future, it would be interesting to calculate the free energy barrier for remodeling, which will give direct kinetic information, at least for a single sequence.

In an effort to observe the impact of the RHP complexity, we examine two-component systems using the same parameterization and protocol as our four-component RHP as points of comparison. Experimental research has investigated poly(MMA-r-OEGMA) in nearly the same composition as ours, and clearly shown that it folds into unimer micelles in water.<sup>51,79</sup> We see that our two-component control sequences become compact as prior theory and experiments predict. Additionally, as evidenced by dihedral angle data in Table 1, the two-component systems are less vitrified than the four-component system and exhibit slightly more backbone motion. Previous theoretical work has investigated RHP freezing temperature as a function of the number of components, and generally aligns with our findings presented here.<sup>14,16,22</sup>

**TABLE 1. Backbone Dihedral Angle Deviations\*\***

Four-component RHP	Poly(MMA-r-OEGMA)	Poly(MMA-r-EHMA)	Poly(MMA-r-SPMA)
50:25:20:5	75:25:0:0	80:0:20:0	95:0:0:5
$1.5 \pm 1.0$ %	$3.4 \pm 1.6$ %	$3.2 \pm 1.4$ %	$5.3 \pm 2.9$ %

\*\*Percentage of the 197 dihedral angles which, over the 40 ns production simulation at 300 K, have a standard deviation of their absolute values greater than 20 Degrees. Mean percentage values with standard deviations are provided for the four-component RHPs and two-component sequences. Sequence compositions are shown as ratios of MMA:OEGMA:EHMA:SPMA. Graphical representation of this data is presented in Supplementary Figure 9.

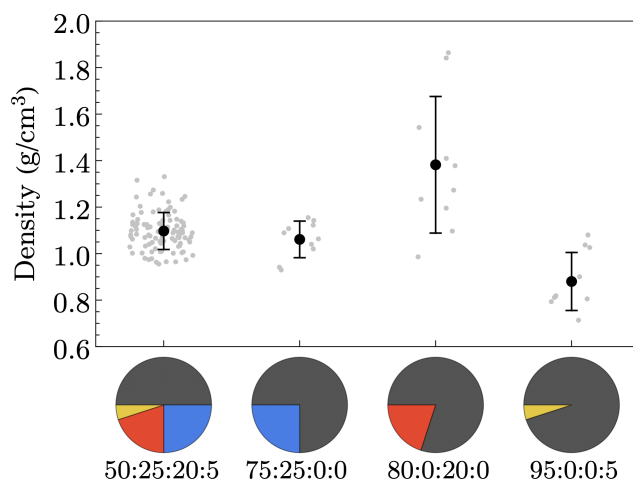


Figure 5. Density of random heteropolymers in water. The density is approximated using convex hull analysis. Individual conformation densities are shown as gray points while overall mean and standard deviation values are indicated in black. Sequence compositions are written as ratios of MMA:OEGMA:EHMA:SPMA.

However, our complex RHP system contains four distinct chemistries, each with unique sizes and degrees of hydrophobicity and hydrophilicity, something not present in the theories. Nevertheless, a simple argument would show that the entropy of mixing within a dense core of the four-component polymer globule is larger than that compared to a similarly dense two-component system, which would point towards the fact that more complex RHPs would fold into denser globules. This presumably would in turn limit backbone mobility, as we see in our results. While this sounds plausible, we have found that density and backbone mobility are not directly correlated. This result is in itself quite important since it decouples the dynamics from the equilibrium aspects, and points toward the effects of specific chemistries on the mobility. This can be seen by comparing the densities of the two-component and four-component systems (Figure 5) and contrasting that with the dihedral rearrangements shown in Table 1. Clearly the poly(MMA-r-EHMA) has a higher density than any of the other systems, yet the rearrangement of the dihedral angles is not as constrained as in the case of a four-component system.

**TABLE 2. Equilibrium Root-Mean-Square Fluctuations<sup>††</sup>**

	Overall	MMA	OEGMA	EHMA	SPMA
Backbone	0.9±0.3 Å	0.9±0.4 Å	0.8±0.3 Å	0.8±0.3 Å	1.0±0.4 Å
Side-Chain	2.7±2.6 Å	1.2±0.5 Å	6.1±3.0 Å	2.0±0.9 Å	2.8±0.7 Å

<sup>††</sup>Mean values with standard deviations provided for the RHPs at 300 K for 40 ns including all monomers (Overall) and broken down by monomer identity. Backbone is measured from  $\alpha$ -carbon atoms while side-chain uses tail atoms, as identified in Supplementary Figure 3.

In contrast to the vitrified backbones in our four-component molten globules, the side-chains show significantly higher mobility at equilibrium. The monomer side-chains experience greater fluctuations than their corresponding backbone atoms, though the degree of change are distinct for each of the four components, as shown in Table 2. The differences in mobilities for each side-chain type indicate that the theory predicting side chain motion in a molten globule to be impacted by available rotational motion and chemical identities remains true with a dynamic solvent environment and the freedom for backbone rearrangement and variable density. OEGMA exhibits the most profoundly mobile side-chain, which corresponds to its long side-chains and unique chemical properties. Our simulations incorporate the solvation and vicinity of neighbors in atomistic detail, which is important for characterizing OEGMA given that PEG, which constitutes the OEGMA side chain, is not modeled well with implicit solvent as its hydrophilic character requires the detail of local water structure and any changes that neighboring atoms may cause to that structure. EHMA and SPMA have similar side-chain lengths to one another, but, interestingly, EHMA exhibits more limited fluctuations. This indicates some hindered motion for EHMA and more free motion of SPMA. We further reveal information about this comparison through observation of the side-chain fluctuations during annealing in Figure 6. At high temperature, the two types of side-chains have nearly equivalent fluctuation magnitudes. As temperature decreases, all fluctuations decrease in an Arrhenius-like fashion, but the rates for EHMA and SPMA differ such that the resulting values are quite distinct. This data indicates that the difference in equilibrium side-chain fluctuations is due to the hydrophobic effect which decreases entropy for EHMA but becomes energetically favorable at low temperature. We can infer from this finding that the EHMA is more likely to be found in the core of the globule, while SPMA is likely at the surface, able to move freely. This observation likens heteropolymer globules to proteins, for which the hydrophobic effect is a dominating factor in their assembly.<sup>80</sup>

It has been experimentally demonstrated by Martini et al. through NMR studies that RHPs with comparable

chemistry (methacrylate-based monomers with PEG and hydrocarbon side-chains) have vitrified backbones and mobile phases made up of the side-chains in the bulk with no sign of large-scale phase segregation.<sup>50</sup> Our findings indicate that this bulk behavior extends to dilute solution for a related system. Guazzelli et al. found trends in SASA for another similar system—using fluorocarbon, rather than hydrocarbon, side-chains as a hydrophobic component—which demonstrated a higher accessibility for the PEG side-chains as compared to the fluorocarbons, indicating hydrophobic interactions leading to buried fluorocarbon chains in the same manner as we hypothesize for our EHMA side-chains based on fluctuation data.<sup>49</sup> Our system includes two additional monomer species as compared to Martini et al. and Guazzelli et al.’s polymers, introducing a greater variety of dipole-dipole, electrostatic, and hydrophobic forces between monomers. We employ a variety of techniques and expand upon the SASA analysis used by Guazzelli et al. to establish the principles governing the assembly of our complex globule.

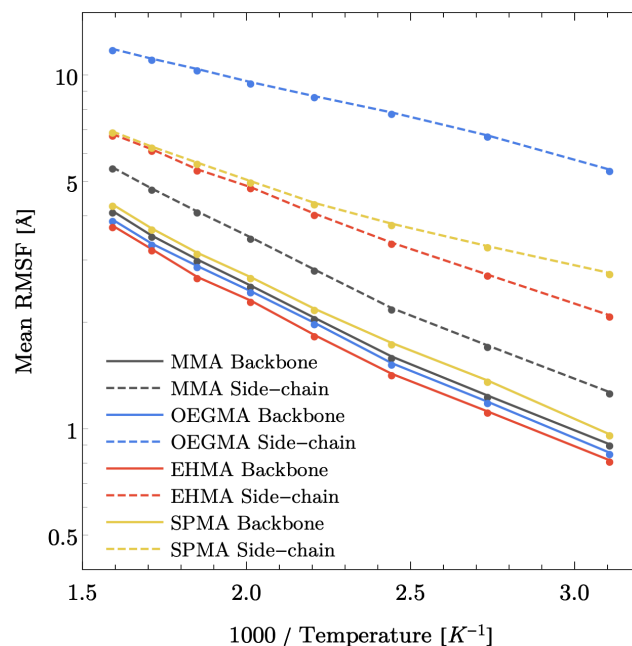


Figure 6. Monomer Mobility. Root-mean-square fluctuations (RMSF) of the backbone (solid lines) and side-chains (dashed lines), as an analog for mobility, plotted as a function of inverse temperature for each monomer type.

In addition to inferences from mobility data, we are able to leverage the backbone vitrification to assess the spatial arrangement of the heteropolymer in solution. At room temperature, dynamical analysis showed that the heteropolymer will not remodel its shape significantly within the nanosecond to microsecond timescale, indicating that average values are appropriate for use in analyzing the 40 ns equilibrated simulations. In our simulations, we anneal at elevated temperature to remodel each sequence and find that the sequences can each form a variety of unique conformations. Our RHPs lack a

sequence-defined structure akin to that of an ordered protein, but have structural trends and conserved features which emerge within and between sequences. Radial distribution functions, visible in Supplementary Figure 7, indicate regular ordering for MMA monomers and disordered packing of SPMA and EHMA monomers in all conformations, showing a dependence on the monomer chemistry to dictate some aspects of RHP spatial arrangement regardless of exact sequence. Despite similar packing trends amongst conformations, each remodeled structure presents distinct dihedral angle and intramolecular distance data. Dihedral angle data akin to Figure 4 for the remaining conformations of sequence 3 are depicted in Supplementary Figure 5 and show unique backbone arrangements for each of the ten runs, as was seen for all sequences. Intramolecular distances, measured between  $\alpha$ -carbon atoms along a sequence and visualized as a contact matrix in which dark clusters (aside

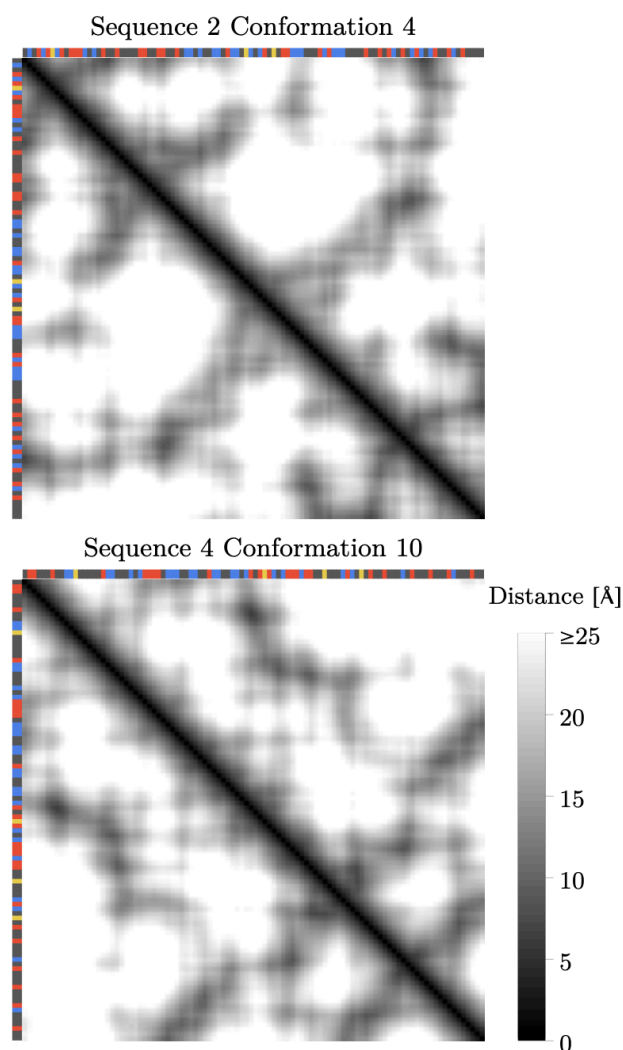


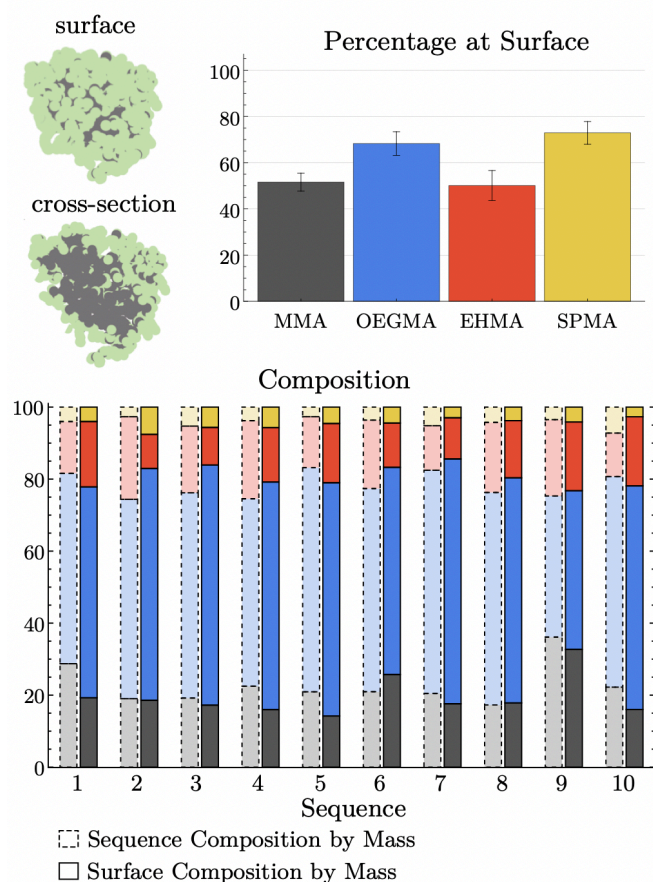
Figure 7. Contact maps from selected heteropolymer conformations. The sequences are colored as indicated in Figure 1 and the distances are calculated between  $\alpha$ -carbon atoms in each monomer. The dark diagonal line corresponds to the polymer backbone given the inherently small distance between nearest neighbors.

from the diagonal created by backbone connectivity) indicate regions of contact in the absence of covalent linkages, also display the uniqueness of each annealed conformation. Several globules show motifs reminiscent of protein structures such as the features parallel, perpendicular, and adjacent to the backbone diagonal, as seen for two selected conformations in Figure 7. These can indicate parallel beta sheets, anti-parallel beta sheets, and alpha helices, respectively, when observed for protein contact maps. While some of the contact maps do show similarities to structural patterns in ordered proteins, the molecular arrangements each represent just one of a multitude of possible low energy states as evidenced by the emergence of several conformations of a given sequence. IDPs can similarly show secondary structure motifs, but the configurations are transient as the driving forces to their folding are not strong enough to create a single energy minimum.<sup>31</sup> The lack of intramolecular hydrogen bonding and racemic nature of our heteropolymers also reduce the driving force for specific secondary structure formation and therefore, the RHPs can adopt multiple distinct conformations for the same sequence.

Analysis thus far highlights regions of different dynamics and ordering. We can then turn to an examination of the globule surface and solvation to reveal interfacial properties and information about potential interactions. The surfaces of the globules are fuzzy due in large part to the long and dynamic PEG chains on OEGMA monomers. Visual inspection shows significant OEGMA presence at the surface as a corona around the immobilized backbone with all four monomer types present at the surface being from time to time shielded by PEG chains, leading to highly heterogeneous shapes and sizes of surface domains. In fact, we believe the transient “hairy” nature of these globules is responsible for their stability in solutions. Snapshots of each conformation are presented in Supplementary Figure 2. The presence of all monomer chemistries at the RHP surfaces, presented graphically in Figure 8 composition plots, is an important property for potential interactions when compared to native proteins, in which approximately half of the accessible area is nonpolar and hydrophobic surface residues play particularly key roles at protein binding interfaces.<sup>80–83</sup> By visual inspection, some angstrom-scale aggregation by monomer type can be appreciated but there are no obvious phase separated microdomains on the RHP surfaces, agreeing with previous theories predicting microphase separation inhibition as a cause for the frozen conformations observed in RHP systems.<sup>18</sup> Rather, the overall surface composition of our system appears to be highly correlated with the sequence composition. This is shown by doing a geometric construction of the surface, and therefore categorizing moieties as surface or non-surface components (Figure 8). This



## A. Surface by Mass



## B. Surface by Monomer

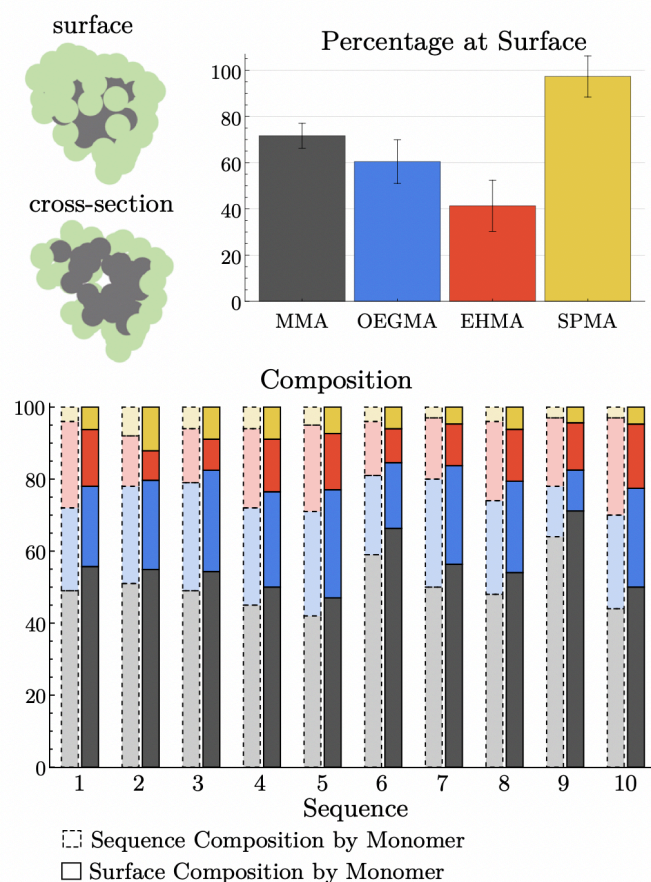


Figure 8. Random Heteropolymer Surface Composition. Percentage of constituents found at the polymer/water interface and sequence breakdown of surface composition (A) by atom and (B) by monomer as found through alpha-shape identification. Data by atom uses all heavy (non-hydrogen) atoms and the resulting compositions are plotted with sequence composition by mass. Monomer data uses the  $C_m$  (Supplementary Figure 3) atom for each residue with sequence composition by number presented for comparison. Compositions for each sequence are the mean values over ten conformations, though individual conformation data can be seen in Supplementary Figure 7. Representative surface and cross-section images shown for each method depict surface constituents in green and core constituents in gray.

is first performed with a single atom per monomer, providing an indication of surface location largely disregarding side-chains. Nearly all SPMA residues are present at the surface due to the hydrophilic and self-repulsive anionic sulfonate group. While EHMA, the most hydrophobic monomer, unsurprisingly has the lowest proclivity to be at the surface, there is intriguingly still a significant portion present with  $41 \pm 11\%$  of residues categorized as surface residues.  $72 \pm 5\%$  of MMA and  $60 \pm 9\%$  of OEGMA monomers are at the surface. It is expected to see these values fall between the propensities of SPMA and EHMA to be present at the surface, but it is counterintuitive that the values imply OEGMA to be less hydrophilic than MMA. However, we recall that the analysis does not account for side-chains and therefore perform another alpha shape identification using all heavy atoms. This dataset maintains the hydrophilicity of SPMA and the hydrophobicity of EHMA, but better depicts more hydrophilic behavior of OEGMA with now  $68 \pm 5\%$  of heavy OEGMA atoms residing at the surface compared to only  $52 \pm 4\%$  of MMA heavy atoms. The differences between these values indicate

that the long OEGMA side-chains contribute significantly to the surface while the carbon backbone is more buried by neighboring monomers or shielded by the side-chain itself. Experimental work has shown that in certain heteropolymer compositions, OEGMA is not restricted to the surface and will have some regions of more limited mobility due to presence in the core, indicating amphiphilicity of OEGMA and aligned with what we observe in simulation.<sup>51,52</sup> For all sequences, it is shown in Composition Plots in Figure 8 that surface composition closely mimics sequence composition. Supplementary Figure 8 shows the composition plots for each individual run and confirms that while the structural conformation of each is unique, the overall composition of the surface is most strongly impacted by the statistical sequence composition and is largely conserved despite geometric differences in chain arrangement.

To confirm findings from alpha shape analysis as well as correlate sequence information to solvation, we calculate the Normalized First Solvation Shell (NFSS), a term we

use to indicate the population of water molecules in the first solvation shell relative to that in a well solvated monomer of the same type. The normalization of the molecular population enables comparison between building blocks of variable chemistry and has been inspired by protein analysis methods.<sup>84</sup> A value of 1 indicates the monomer is as freely accessible as it can be in a chain while a monomer without any water present in the first solvation shell would have an NFSS of 0. NFSS is calculated using all atoms, including hydrogen, for each monomer excluding 5 monomers at each chain end.

Solvation trends observed are consistent with alpha shape surface identification, with mean NFSS value ordering corresponding to the alpha shape results by mass. Mean values of MMA and OEGMA are extremely close while SPMA shows much higher NFSS and EHMA shows much lower NFSS. This result emphasizes the amphiphilicity of both MMA and OEGMA. Furthermore, the sensitivity of these monomers' hydrophilicity and hydrophobicity to analysis technique highlights the complexity of their behavior in water.

The solvation of ordered protein residues can often be predicted through sequence-to-structure relationships. To probe similar correlations in our polymers, we

examine the NFSS of our polymers as a function of first nearest neighbor identities relative to the aforementioned mean values (Figure 9). While the mean value, shown as horizontal lines in Figure 9, is distinct for each monomer type, the deviations from that mean as a function of nearest neighbor (shown by the solid points with standard deviation error bars) trend similarly to one another. When OEGMA is a monomer's nearest neighbor, the NFSS value decreases, likely due to a shielding effect from the long OEGMA side-chains. When both nearest neighbors are OEGMA, MMA and EHMA monomers have significantly lower solvation values. All monomers also have reduced solvation with one OEGMA and one EHMA neighbor. In contrast to OEGMA, SPMA increases solvation for its neighbors due to the anionic sulfonate end group pulling the monomer out into solvent and away from the non-polar moieties. MMA also increases NFSS of its neighbors, but for different reasons. The MMA monomer itself is amphiphilic and does not have a strong driving force to drag its neighbors out into solvent like the SPMA does, but instead impacts the steric interactions. The smallest of the four constituents, MMA opens up the accessibility of its neighbors and is unable to cover hydrophobic regions with polar side-chains in the manner that OEGMA can.

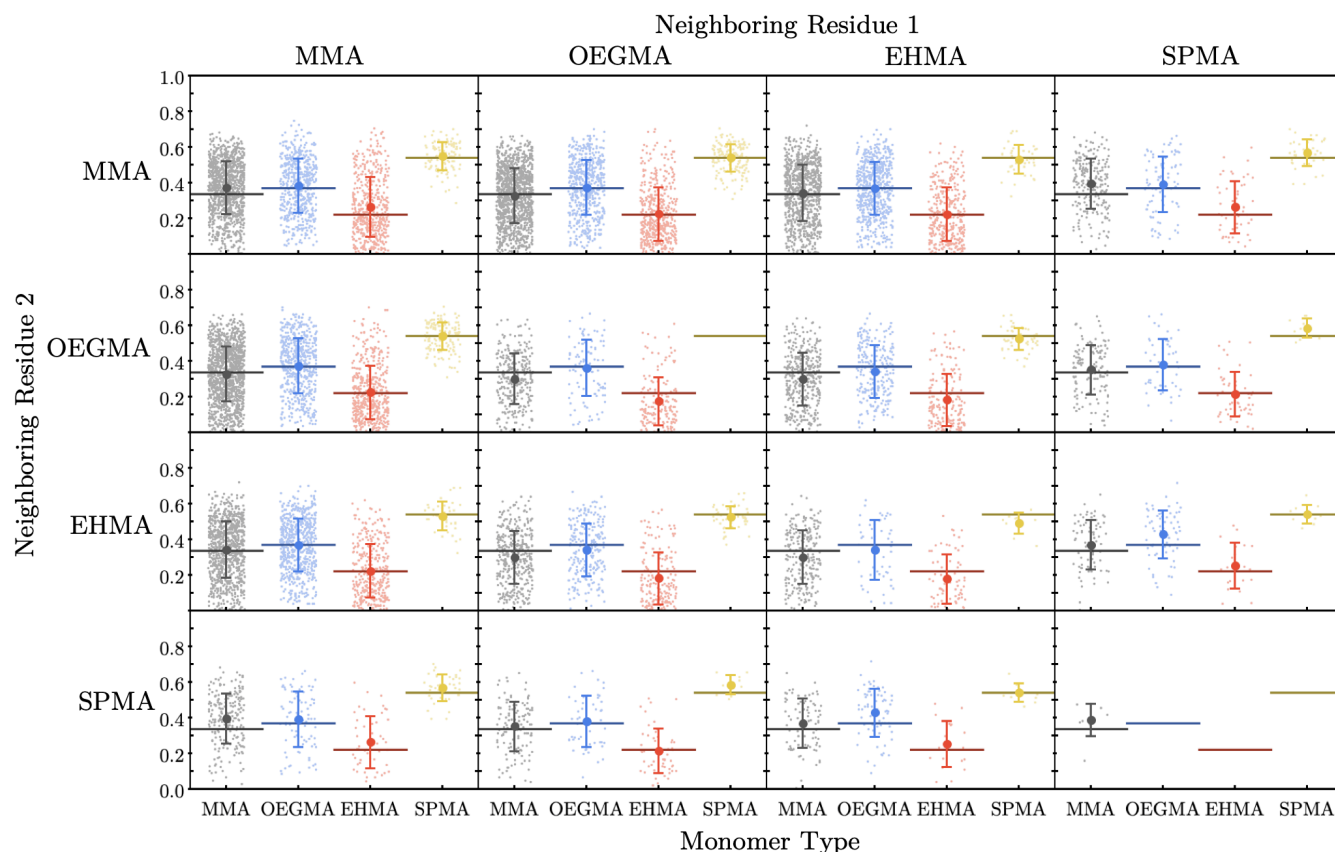


Figure 9. Normalized First Solvation Shell. NFSS data by monomer type and nearest neighbor identities. Overall solvation averages for each monomer type are indicated by horizontal lines in each plot and error bars indicate standard deviation values for a given set of nearest neighbors.

Similar analysis to NFSS gives us a Relative Solvent Accessibility (RSA) based on normalized SASA data. Using this method common in the protein analysis field rather than SASA values more typical of synthetic homopolymer analysis, data can be directly compared for various monomer types in the same manner as our NFSS values. Resulting trends of nearest neighbor investigation largely mirrored the conclusions of NFSS analysis and can be seen in Supplementary Figure 8. RSA relies on algorithms which make approximations, while NFSS uses atomistic information about explicit water molecule location throughout the simulation, and is therefore more reliable for our investigation.

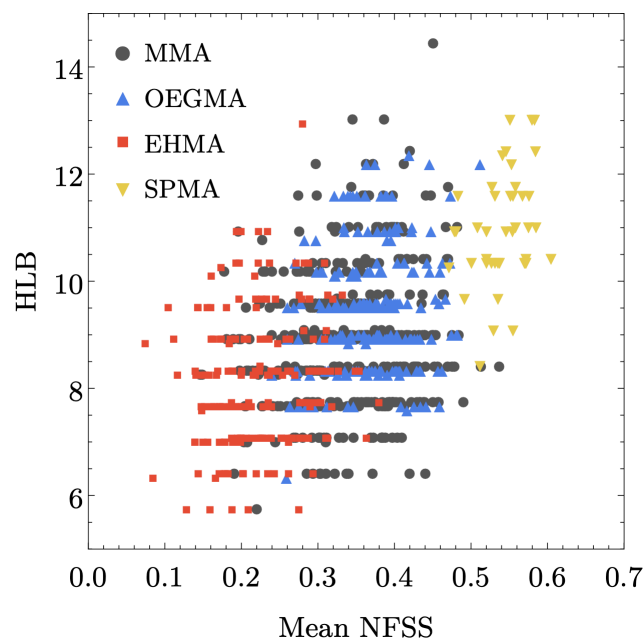


Figure 10. Hydrophilic-Lipophilic Balance versus Normalized First Solvation Shell. Calculated HLB versus NFSS such that each point represents a monomer in one of the ten selected sequences for simulation. The plotted NFSS values are the mean values for each monomer from a given sequence's ten conformations. The data shows a very weak positive correlation between the two measures.

Predictive capability for monomer solvation is highly desirable for our heteropolymers. We can compare our NFSS results to sequence hydrophilic-lipophilic balance (HLB) classifications which attempt to evaluate the solubility of monomer side-chains via group contribution theory.<sup>85</sup> The HLB value of each monomer along a chain is reassigned using an average HLB of a five-monomer window including the monomer itself and its two nearest neighbors on either side with equal weight.<sup>8</sup> However, as we have shown through our MD results, there are steric and chemical factors that impact the hydration. In addition to the sequence connectivity, monomer identity and size are profoundly important in assessing hydration and location within the assembled globule. The exclusion of these factors and the equal weighting of all monomers within the HLB window lead to low correlation between HLB and NFSS on a monomer-by-monomer basis seen

in Figure 10. EHMA and SPMA exhibit behavior consistent with hydrophobic and hydrophilic classifications, respectively. MMA and OEGMA, however, cannot be simply described by these characterizations and demonstrate amphiphilicity. It is convenient to think of each side-chain in terms of its typical solvation character in water as a homopolymer, but our results prove that this can be misleading. Firstly, the connectivity from the backbone must be considered. The RHP connectivity can force chemically disparate regions close together and also, in regions with large populations of massive monomers, the constituents experience crowding akin to bottle-brush polymers which restrict accessibility and can even have local densities greater than solvation limits despite generally hydrophilic properties. Secondly, nanodomains of aggregated hydrocarbon chains, such as those that may form from the EHMA side-chains and polymer backbone, will have reduced local dielectric constants. This will have important ramifications for the interaction strengths of the ester groups and polar side-chains if embedded in a low dielectric region. Our analysis of the molecular interfaces of the RHPs indicate these complex interactions are pivotal in how the amphiphilic molecules behave.

## CONCLUSION

We have explored the structure and dynamics of synthetic RHPs at an atomistic level using simulations. Our results provide strong supporting evidence for previous experimental data and theories, and at the same time give access to very fine molecular information that is difficult, if not impossible, to obtain experimentally. By simulating individual sequences, we isolated distinct components and sequence dependent properties, bridging the gap between atomic and collective properties obtained when analyzing polymers through experiments.<sup>53</sup> Furthermore, our work provides a framework for simulating real, synthesizable materials and producing results that can augment and inform experimental techniques. We observed that the RHP globules investigated in this work are stochastic, without a predetermined structure. However, the collapse and assembly seem to be controlled in an analogous fashion to proteins, maintaining the general features predicted from mean field models of RHP folding.<sup>24</sup> Additionally, the interfaces of the single chain particles universally present heterogeneous patchy surfaces, similar to proteins. The collapsed RHP globules with vitrified backbones and conformational patterns akin to secondary structure draw further comparison to the ordered biomolecules, yet the fact that RHPs display multiple stable conformations places them somewhere in between native ordered and intrinsically disordered proteins. In the future, controlling this dynamicity by changing chemistries might open avenues for different functionalities, as we know is the case for commonly used plastics. Analysis of the heteropolymer globule shows a dependence upon chemical identity, with

regards to both steric effects and polarity, as well as upon molecular connectivity, showing that the chemical complexity impacts the morphology and dynamics of RHPs. In particular, OEGMA is frequently used in SCNP research and our exposition of how it behaves in the RHP construct here is important for future works exploiting this chemistry, especially because this monomer cannot be well described by simplified analytical models due to its dual hydrophobic-hydrophilic character dependent on the local environment. While proteins do not have components as large and complex as OEGMA, it would be interesting to extend these findings to the effect that post-translational modifications may have in proteins. For example, natural peptides can undergo modifications such as glycosylation, resulting in large and bulky side groups that could in principle similarly display a dual hydrophobic-hydrophilic nature. The impact of these large moieties could be informed by our studies and become relevant in protein work investigating biological phenomena such as liquid-liquid phase separation and protein folding, misfolding, and oligomerization. Overall, our work provided the first atomic level characterization of complex RHP globules (or SCNPs) and will provide the basis to understand the function of such macromolecules, including but not limited to observed inter-globular interactions. We can anticipate that due to the rich energy landscape created by the heterogenous surfaces of RHPs, we expect them to display interesting behaviors as it pertains to their interactions with natural proteins, lipids, and other molecules, as recent experimental work has observed.<sup>6,8</sup> Thus, we envision this work will open new possibilities to understand and design novel functionalities for RHPs.

## ASSOCIATED CONTENT

### Supporting Information

Overall vs. selected sequence distribution plot, snapshots from simulations, atom identification for analysis, additional SAXS data, additional dihedral angle plots, Radial Distribution Function (RDF) plots, Surface composition for each conformation, Relative Solvent Accessibility (RSA) plots (PDF)

## AUTHOR INFORMATION

### Corresponding Author

\* aalexand@mit.edu

### Funding Sources

This work was supported by the Defense Threat Reduction Agency contract HDTRA11910011.

### Conflicts of Interest

There are no conflicts of interest to declare.

## REFERENCES

- (1) Shakhnovich, E. I.; Gutin, A. M. Frozen states of a disordered globular heteropolymer. *J. Phys. A: Math. Gen.* **1989**, 22 (10), 1647–1659.
- (2) Maier, C.; Calafut, T. *Polypropylene: The Definitive User's Guide and Databook*. Polypropylene: The Definitive User's Guide and Databook; Plastics Design Library: Norwich, NY,

- 1998.
- (3) Rudnick, L. R.; Shubkin, R. L. *Synthetic Lubricants and High-Performance Functional Fluids*. *Synthetic Lubricants and High-Performance Functional Fluids*, 2nd ed.; Marcel Dekker, Inc.: New York, NY, 1999.
- (4) Huang, E.; Russell, T. P.; Harrison, C.; Chaikin, P. M.; Register, R. A.; Hawker, C. J.; Mays, J. *Using surface active random copolymers to control the domain orientation in diblock copolymer thin films*. *Using surface active random copolymers to control the domain orientation in diblock copolymer thin films*; 1998; Vol. 31.
- (5) Huurme, G. M. ter; Palmans, A. R. A.; Meijer, E. W. Supramolecular Single-Chain Polymeric Nanoparticles. *CCS Chem.* **2019**, 64–82.
- (6) Panganiban, B.; Qiao, B.; Jiang, T.; DelRe, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A. A.; Hall, A.; Sit, I.; Crosby, M. G.; Dennis, P. B.; Drockenmüller, E.; Olvera de la Cruz, M.; Xu, T. Random heteropolymers preserve protein function in foreign environments. *Science (80-. )*. **2018**, 359 (6381), 1239–1243.
- (7) Nguyen, T. D.; Qiao, B.; Olvera de la Cruz, M. Efficient encapsulation of proteins with random copolymers. *Proc. Natl. Acad. Sci.* **2018**, 115 (26), 6578–6583.
- (8) Jiang, T.; Hall, A.; Eres, M.; Hemmatian, Z.; Qiao, B.; Zhou, Y.; Ruan, Z.; Couse, A. D.; Heller, W. T.; Huang, H.; de la Cruz, M. O.; Rolandi, M.; Xu, T. Single-chain heteropolymers transport protons selectively and rapidly. *Nature* **2020**, 577 (7789), 216–220.
- (9) Shakhnovich, E. I.; Gutin, A. M. Formation of unique structure in polypeptide chains. *Biophys. Chem.* **1989**, 34 (3), 187–199.
- (10) Pande, V. S.; Rokhsar, D. S. Is the molten globule a third phase of proteins? *Proc. Natl. Acad. Sci. U. S. A.* **1998**, 95 (4), 1490–1494.
- (11) Lau, K. F.; Dill, K. A. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules* **1989**, 22 (10), 3986–3997.
- (12) Ramanathan, S.; Shakhnovich, E. Statistical mechanics of proteins with “evolutionary selected” sequences. *Phys. Rev. E* **1994**, 50 (2), 1303–1312.
- (13) Chakraborty, A. K.; Shakhnovich, E. I.; Pande, V. S. Freezing of compact random heteropolymers with correlated sequence fluctuations. *J. Chem. Phys.* **1998**, 108 (4), 1683–1687.
- (14) Sfatos, C. D.; Gutin, A. M.; Shakhnovich, E. I. Phase diagram of random copolymers. *Phys. Rev. E* **1993**, 48 (1), 465–475.
- (15) Shakhnovich, E.; Gutin, A. Enumeration of all compact conformations of copolymers with random sequence of links. *J. Chem. Phys.* **1990**, 93 (8), 5967–5971.
- (16) Pande, V. S.; Grosberg, A. Y.; Tanaka, T. Freezing transition of random heteropolymers consisting of an arbitrary set of monomers. *Phys. Rev. E* **1995**, 51 (4), 3381–3392.
- (17) Dima, R. I.; Banavar, J. R.; Cieplak, M.; Maritan, A. Statistical mechanics of protein-like heteropolymers. *Proc. Natl. Acad. Sci.* **1999**, 96 (9), 4904–4907.
- (18) Sfatos, C. D.; Gutin, A. M.; Shakhnovich, E. I. Phase transitions in a many-letter random heteropolymer. *Phys. Rev. E* **1994**, 50 (4), 2898–2905.
- (19) Davidson, A. R.; Lumb, K. J.; Sauer, R. T. Cooperatively folded proteins in random sequence libraries. *Nat. Struct. Biol.* **1995**, 2 (10), 856–864.
- (20) Hua, Q. X.; Ladbury, J. E.; Weiss, M. A. Dynamics of a monomeric insulin analog: Testing the molten-globule hypothesis. *Biochemistry* **1993**, 32 (6), 1433–1442.
- (21) Peng, Z. Y.; Kim, P. S. A Protein Dissection Study of a Molten Globule. *Biochemistry* **1994**, 33 (8), 2136–2141.
- (22) Shakhnovich, E. Protein Folding Thermodynamics and Dynamics: Where Physics, Chemistry, and Biology Meet. *Chem. Rev.* **2006**, 106 (5), 1559–1588.
- (23) Dill, K. A.; Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Mol. Biol.* **1997**, 4 (1), 10–19.
- (24) Shakhnovich, E. I.; Gutin, A. M. Implications of thermodynamics of protein folding for evolution of primary sequences. *Nature* **1990**, 346 (6286), 773–775.
- (25) Riddle, D. S.; Santiago, J. V.; Bray-Hall, S. T.; Doshi, N.; Grantcharova, V. P.; Yi, Q.; Baker, D. Functional rapidly folding proteins from simplified amino acid sequences. *Nat. Struct. Biol.* **1997**, 4 (10), 805–809.



- (26) Fan, K.; Wang, W. What is the Minimum Number of Letters Required to Fold a Protein? *J. Mol. Biol.* **2003**, *328* (4), 921–926.
- (27) Pande, V. S.; Grosberg, A. Y. Heteropolymer freezing and design: Towards physical models of protein folding. *Phase Transitions* **2000**, *72* (1), 259–314.
- (28) Ma, C. D.; Wang, C.; Acevedo-Vélez, C.; Gellman, S. H.; Abbott, N. L. Modulation of hydrophobic interactions by proximally immobilized ions. *Nature* **2015**, *517* (7534), 347–350.
- (29) Wang, C.; Ma, C. K. D.; Yeon, H.; Wang, X.; Gellman, S. H.; Abbott, N. L. Nonadditive Interactions Mediated by Water at Chemically Heterogeneous Surfaces: Nonionic Polar Groups and Hydrophobic Interactions. *J. Am. Chem. Soc.* **2017**, *139* (51), 18536–18544.
- (30) Salvi, N.; Abyzov, A.; Blackledge, M. Analytical Description of NMR Relaxation Highlights Correlated Dynamics in Intrinsically Disordered Proteins. *Angew. Chemie Int. Ed.* **2017**, *56* (45), 14020–14024.
- (31) Uversky, V. N. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Front. Phys.* **2019**, *7* (FEB), 8–23.
- (32) Uversky, V. N. A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Sci.* **2013**, *22* (6), 693–724.
- (33) Kulkarni, P.; Uversky, V. N. Intrinsically Disordered Proteins: The Dark Horse of the Dark Proteome. *Proteomics* **2018**, *18* (21–22), 1800061.
- (34) Larion, M.; Miller, B.; Brüscheweiler, R. Conformational heterogeneity and intrinsic disorder in enzyme regulation: Glucokinase as a case study. *Intrinsically Disord. Proteins* **2015**, *3* (1), e1011008.
- (35) Schuler, B.; Hofmann, H. Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales. *Curr. Opin. Struct. Biol.* **2013**, *23* (1), 36–47.
- (36) Chen, J.; Kriwacki, R. W. Intrinsically Disordered Proteins: Structure, Function and Therapeutics. *J. Mol. Biol.* **2018**, *430* (16), 2275–2277.
- (37) Ter Huurne, G. M.; Voets, I. K.; Palmans, A. R. A.; Meijer, E. W. Effect of Intra- versus Intermolecular Cross-Linking on the Supramolecular Folding of a Polymer Chain. *Macromolecules* **2018**, *51* (21), 8853–8861.
- (38) Hanlon, A. M.; Lyon, C. K.; Berda, E. B. What Is Next in Single-Chain Nanoparticles? *Macromolecules* **2016**, *49* (1), 2–14.
- (39) Oberdisse, J.; González-Burgos, M.; Mendia, A.; Arbe, A.; Moreno, A. J.; Pomposo, J. A.; Radulescu, A.; Colmenero, J. Effect of Molecular Crowding on Conformation and Interactions of Single-Chain Nanoparticles. *Macromolecules* **2019**, *52* (11), 4295–4305.
- (40) Rubio-Cervilla, J.; González, E.; Pomposo, J. Advances in Single-Chain Nanoparticles for Catalysis Applications. *Nanomaterials* **2017**, *7* (10), 341.
- (41) Zhu, J.; Han, H.; Li, F.; Wang, X.; Yu, J.; Chu, C.-C.; Wu, D. Self-assembly of amino acid-based random copolymers for antibacterial application and infection treatment as nanocarriers. *J. Colloid Interface Sci.* **2019**, *540*, 634–646.
- (42) Ko, J. H.; Bhattacharya, A.; Terashima, T.; Sawamoto, M.; Maynard, H. D. Amphiphilic fluorinated random copolymer self-assembly for encapsulation of a fluorinated agrochemical. *J. Polym. Sci. Part A Polym. Chem.* **2019**, *57* (3), 352–359.
- (43) Sanchez-Sanchez, A.; Akbari, S.; Etxeberria, A.; Arbe, A.; Gasser, U.; Moreno, A. J.; Colmenero, J.; Pomposo, J. A. “Michael” Nanocarriers Mimicking Transient-Binding Disordered Proteins. *ACS Macro Lett.* **2013**, *2* (6), 491–495.
- (44) Liu, Y.; Pujals, S.; Stals, P. J. M.; Paulöhr, T.; Presolski, S. I.; Meijer, E. W.; Albertazzi, L.; Palmans, A. R. A. Catalytically Active Single-Chain Polymeric Nanoparticles: Exploring Their Functions in Complex Biological Media. *J. Am. Chem. Soc.* **2018**, *140* (9), 3423–3433.
- (45) Artar, M.; Souren, E. R. J.; Terashima, T.; Meijer, E. W.; Palmans, A. R. A. Single Chain Polymeric Nanoparticles as Selective Hydrophobic Reaction Spaces in Water. *ACS Macro Lett.* **2015**, *4* (10), 1099–1103.
- (46) Tooley, C. A.; Pazicni, S.; Berda, E. B. Toward a tunable synthetic [FeFe] hydrogenase mimic: Single-chain nanoparticles functionalized with a single diiron cluster. *Polym. Chem.* **2015**, *6* (44), 7646–7651.
- (47) Cole, J. P.; Hanlon, A. M.; Rodriguez, K. J.; Berda, E. B. Protein-like structure and activity in synthetic polymers. *J. Polym. Sci. Part A Polym. Chem.* **2017**, *55* (2), 191–206.
- (48) Upadhyaya, R.; Murthy, N. S.; Hoop, C. L.; Kosuri, S.; Nanda, V.; Kohn, J.; Baum, J.; Gormley, A. J. PET-RAFT and SAXS: High Throughput Tools To Study Compactness and Flexibility of Single-Chain Polymer Nanoparticles. *Macromolecules* **2019**, *52* (21), 8295–8304.
- (49) Guazzelli, E.; Martinelli, E.; Galli, G.; Cupellini, L.; Jurinovich, S.; Mennucci, B. Single-chain self-folding in an amphiphilic copolymer: An integrated experimental and computational study. *Polymer (Guildf.)* **2019**, *161* (November 2018), 33–40.
- (50) Martini, F.; Guazzelli, E.; Martinelli, E.; Borsacchi, S.; Geppi, M.; Galli, G. Molecular Dynamics of Amphiphilic Random Copolymers in the Bulk: A <sup>1</sup>H and <sup>19</sup>F NMR Relaxometry Study. *Macromol. Chem. Phys.* **2019**, *220* (16), 1900177.
- (51) Imai, S.; Hirai, Y.; Nagao, C.; Sawamoto, M.; Terashima, T. Programmed Self-Assembly Systems of Amphiphilic Random Copolymers into Size-Controlled and Thermoresponsive Micelles in Water. *Macromolecules* **2018**, *51* (2), 398–409.
- (52) Shibata, M.; Matsumoto, M.; Hirai, Y.; Takenaka, M.; Sawamoto, M.; Terashima, T. Intramolecular Folding or Intermolecular Self-Assembly of Amphiphilic Random Copolymers: On-Demand Control by Pendant Design. *Macromolecules* **2018**, *51* (10), 3738–3745.
- (53) Gonzalez-Burgos, M.; Latorre-Sanchez, A.; Pomposo, J. A. Advances in single chain technology. *Chem. Soc. Rev.* **2015**, *44* (17), 6122–6142.
- (54) Smith, A. A. A.; Hall, A.; Wu, V.; Xu, T. Practical Prediction of Heteropolymer Composition and Drift. *ACS Macro Lett.* **2019**, *8* (1), 36–40.
- (55) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revision D.01. Gaussian, Inc.: Wallingford, CT 2004.
- (56) Case, D. A.; Cerutti, D. S.; Cheatham, T. E. I.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, J.; Liu, T.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. AMBER 2017. University of California: San Francisco 2017.
- (57) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112* (30), 9020–9041.
- (58) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Jarrod Millman, K.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, I.; Ilhan, F.; Moore, E. W.; Vand erPlas, J.; Laxalde, D.; Perktold, J.; Cimman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Contributors, S. I. 0. SciPy 1.0—Fundamental Algorithms for Scientific Computing in Python. *arXiv e-prints* **2019**, arXiv:1907.10121.
- (59) van der Walt, S.; Colbert, S. C.; Varoquaux, G. The NumPy

- Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13* (2), 22–30.
- (60) Albou, L.-P.; Schwarz, B.; Poch, O.; Wurtz, J. M.; Moras, D. Defining and characterizing protein surface using alpha shapes. *Proteins Struct. Funct. Bioinforma.* **2009**, *76* (1), 1–12.
- (61) Zhou, W.; Yan, H. Alpha shape and Delaunay triangulation in studies of protein-related interactions. *Brief. Bioinform.* **2014**, *15* (1), 54–64.
- (62) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38.
- (63) Danilov, D.; Barner-Kowollik, C.; Wenzel, W. Modelling of reversible single chain polymer self-assembly: From the polymer towards the protein limit. *Chem. Commun.* **2015**, *51* (27), 6002–6005.
- (64) Salvi, G.; De Los Rios, P. Effective Interactions Cannot Replace Solvent Effects in a Lattice Model of Proteins. *Phys. Rev. Lett.* **2003**, *91* (25), 2581021–2581024.
- (65) Daura, X.; Mark, A. E.; Van Gunsteren, W. F. Peptide folding simulations: No solvent required? *Comput. Phys. Commun.* **1999**, *123* (1–3), 97–102.
- (66) Receveur-Brechot, V.; Durand, D. How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective. *Curr. Protein Pept. Sci.* **2012**, *13* (1), 55–75.
- (67) Pollack, L.; Tate, M. W.; Darnton, N. C.; Knight, J. B.; Gruner, S. M.; Eaton, W. A.; Austin, R. H. Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle x-ray scattering. *Proc. Natl. Acad. Sci.* **1999**, *96* (18), 10115–10117.
- (68) Anufrieva, E. V.; Bychkova, V. E.; Krakovyak, M. G.; Pautov, V. D.; Ptitsyn, O. B. A synthetic polypeptide with a compact structure and its self-organization. *FEBS Lett.* **1975**, *55* (1–2), 46–49.
- (69) Guo, Y.; Morozov, A.; Schneider, D.; Chung, J. W.; Zhang, C.; Waldmann, M.; Yao, N.; Fytas, G.; Arnold, C. B.; Priestley, R. D. Ultrastable nanostructured polymer glasses. *Nat. Mater.* **2012**, *11* (4), 337–343.
- (70) Liberman, S. A.; Gomes, A. D. S.; Macchi, E. M. Compatibility in poly(ethylene oxide)–poly(methyl methacrylate) blends. *J. Polym. Sci. Polym. Chem. Ed.* **1984**, *22* (11), 2809–2815.
- (71) Malcolm, G. N.; Rowlinson, J. S. The thermodynamic properties of aqueous solutions of polyethylene glycol, polypropylene glycol and dioxane. *Trans. Faraday Soc.* **1957**, *53*, 921.
- (72) Finkelstein, A. V.; Shakhnovich, E. I. Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers* **1989**, *28* (10), 1681–1694.
- (73) Baldwin, R. L.; Rose, G. D. Molten globules, entropy-driven conformational change and protein folding. *Curr. Opin. Struct. Biol.* **2013**, *23* (1), 4–10.
- (74) Ptitsyn, O. B. In *Journal of Industrial and Engineering Chemistry*; 1995; pp 83–229.
- (75) Shakhnovich, E. I.; Finkelstein, A. V. Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers* **1989**, *28* (10), 1667–1680.
- (76) Salvi, N.; Abyzov, A.; Blackledge, M. Multi-Timescale Dynamics in Intrinsically Disordered Proteins from NMR Relaxation and Molecular Simulation. *J. Phys. Chem. Lett.* **2016**, *7* (13), 2483–2489.
- (77) Rezaei-Ghaleh, N.; Parigi, G.; Soranno, A.; Holla, A.; Becker, S.; Schuler, B.; Luchinat, C.; Zweckstetter, M. Local and Global Dynamics in Intrinsically Disordered Synuclein. *Angew. Chemie Int. Ed.* **2018**, *57* (46), 15262–15266.
- (78) Gallat, F.-X.; Laganowsky, A.; Wood, K.; Gabel, F.; van Eijck, L.; Wuttke, J.; Moulin, M.; Härtlein, M.; Eisenberg, D.; Colletier, J.-P.; Zaccari, G.; Weik, M. Dynamical Coupling of Intrinsically Disordered Proteins and Their Hydration Water: Comparison with Folded Soluble and Membrane Proteins. *Biophys. J.* **2012**, *103* (1), 129–136.
- (79) Shibata, M.; Matsumoto, M.; Hirai, Y.; Takenaka, M.; Sawamoto, M.; Terashima, T. Intramolecular Folding or Intermolecular Self-Assembly of Amphiphilic Random Copolymers: On-Demand Control by Pendant Design. *Macromolecules* **2018**, *51* (10), 3738–3745.
- (80) Tsai, C.-J.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Sci.* **2008**, *6* (1), 53–64.
- (81) Jones, S.; Thornton, J. M. Analysis of protein-protein interaction sites using surface patches 1 Edited by G.Von Heijne. *J. Mol. Biol.* **1997**, *272* (1), 121–132.
- (82) Rose, G.; Geselowitz, A.; Lesser, G.; Lee, R.; Zehfus, M. Hydrophobicity of amino acid residues in globular proteins. *Science* (80-. ). **1985**, *229* (4716), 834–838.
- (83) Gerstein, M.; Chothia, C. Packing at the protein-water interface. *Proc. Natl. Acad. Sci.* **1996**, *93* (19), 10167–10172.
- (84) Chothia, C. The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **1976**, *105* (1), 1–12.
- (85) Davies, J. T. A Quantitative Kinetic Theory of Emulsion Type. I. Physical Chemistry of the Emulsifying Agent. In *Proc. 2nd Int. Congr. Surface Activity*, vol. 1, London; 1957; pp 426–438.